Exploring Metagenomics in the Laboratory of an Introductory Biology Course†

Brian B. Gibbens††††, Cheryl L. Scott††††, Courtney D. Hoff†, and Janet L. Schottel²

1Department of Biology Teaching and Learning, 2Department of Biochemistry, Molecular Biology & Biophysics, University of Minnesota, St. Paul, MN 55108

Four laboratory modules were designed for introductory biology students to explore the field of metagenomics. Students collected microbes from environmental samples, extracted the DNA, and amplified 16S rRNA gene sequences using polymerase chain reaction (PCR). Students designed functional metagenomics screens to determine and compare antibiotic resistance profiles among the samples. Bioinformatics tools were used to generate and interpret phylogenetic trees and identify homologous genes. A pretest and posttest were used to assess learning gains, and the results indicated that these modules increased student performance by an average of 22%. Here we describe ways to engage students in metagenomics-related research and provide readers with ideas for how they can start developing metagenomics exercises for their own classrooms.

INTRODUCTION

Metagenomics is a culture-independent technique that strives to collect and analyze the complete microbial DNA contained in a given environmental sample. The rapidly advancing field of metagenomics promises real-world applications and advances for medicine, industry, alternative energy, and environmental remediation (1). Given estimates that less than 1% of the microbes in an environmental sample can be cultured in the laboratory (4, 7), metagenomics has grown to be an important tool for analyzing populations of organisms for taxonomic diversity and as a potentially rich source of novel proteins and enzymes (8). As a highly multidisciplinary field, metagenomics makes an ideal teaching platform to convey basic concepts of microbial diversity and evolution (2). The straightforward techniques and tools used in metagenomics can be successfully used by relatively inexperienced students and have the potential to generate vast amounts of mineable data. Students will experience the research process by developing hypotheses, designing experiments, and analyzing data they generate in the laboratory. While learning basic principles, students have an opportunity to contribute to the rapidly expanding field of microbial ecology and experience first-hand the excitement of scientific discovery (3).

The metagenomic modules we developed make use of the Minnesota Mississippi Metagenomics Project (M3P) (http://www.mississippi-metagenome-project.umn.edu/), which is an ongoing project that examines the microbial populations in different regions of the Mississippi River extending from its origin at Lake Itasca to its confluence with the Zumbro River. Sampling sites differ in agricultural use, population density, industrial development and potential sources of contamination, which can influence the types of microbial populations found in the local land and water habitats (5, 6). Although the laboratory modules we have designed take advantage of the resources available from the M3P, environmental samples can be taken from many different sources, and the modules can be easily adjusted to accommodate different geographical locations or varieties of ecosystems.

Here we describe four laboratory modules that use sequence-based and function-based metagenomics as a platform for students to explore microbial populations in environmental samples and to formulate hypotheses regarding changes in microbial populations in response to different environmental conditions. The modules include collecting environmental microbes, extracting DNA from those microbes, designing a functional metagenomics screen, and analyzing metagenomic sequence data using online bioinformatics tools. The laboratory modules were created to introduce students to the theory and applications of metagenomics while simultaneously addressing several different basic biology concepts.

Intended audience

The metagenomic modules were designed for students in an introductory biology course who are non-science

††Brian Gibbens and Cheryl Scott contributed equally to this manuscript.
majors, early in their college careers, with no prior laboratory experience. These modules would be equally appropriate for use in introductory courses for students who intend to major in Biology or related disciplines such as Microbiology, Genetics, or Biotechnology.

Learning time

Biology 1009 has both lecture and laboratory components. The lecture meets for 75 minutes twice a week, and the laboratory section of the course meets once a week for two hours. Each laboratory meeting usually begins with a brief quiz to assess student preparedness for the lab, followed by a 5–10 minute introduction to the lab activity by the teaching assistant. The introduction emphasizes the learning objectives and major concepts for the lab, general procedures, and safety issues. Each of the metagenomic modules takes approximately 100 minutes to complete.

Prerequisite student knowledge

Most students in the class will have had a biology course in high school and perhaps some preparation in chemistry. However, there are no college-level prerequisites for this course and no prior laboratory experience is expected or required. The laboratory portion of the course is taken in conjunction with a lecture. The subjects covered in lecture generally coincide with the order of the laboratory modules, and thus the lecture material provides substantial background information for the lab. The coordination between the lecture and laboratory limits the placement of the metagenomic modules to particular times during the semester. The textbook, lab manual, and other resources provide students with additional background information.

Learning objectives

Upon completion of these modules, students will be able to:

1. Define metagenomics and explain how it is used to study microbial populations in environmental samples.
2. Explain rarefaction curves and operational taxonomic units and how 16S rRNA gene sequences can differ among organisms in an environmental sample.
3. Compare sequence-based and function-based metagenomics; explain how 16S rRNA gene sequences can classify microbes and how a fosmid library of genomic sequences can be used to identify novel proteins.
4. Use bioinformatic tools to classify organisms, develop a phylogenetic tree, and identify homologous genes.

In addition to these overarching learning objectives, each of the four laboratory modules has several specific learning objectives (Table 1).

PROCEDURE

Materials and equipment

Four metagenomics lab modules were performed during the course of the semester: “The Microbial World,” “Extracting the Microbial Metagenome,” “Sequence- and Function-Based Metagenomics,” and “Bioinformatic Analysis of Metagenomes.” The four labs were designed to be carried out in succession; however, this is not a requirement (see possible modifications section). To complete each lab module, the following materials, reagents and equipment were used (a complete and detailed list is provided in Appendix 1).

The Microbial World
• Glass slides and coverslips
• Immersion oil
• Phosphate buffered saline (PBS)
• 0.45-μm cellulose filter
• 5.0-μm cellulose filter
• Filter holder
• 50-mL syringes
• *Metal forceps
• *1.5-mL microcentrifuge tubes
• 15-mL conical tubes
• *Micropipettes
• Light microscope
• Centrifuges (to accommodate 1.5-mL microcentrifuge and 15-mL conical tubes)
• Methyl blue and methyl green stains
• Commercially prepared bacterial slides
• Live bacterial and eukaryotic cultures

*Also needed for Extracting the Microbial Metagenome and Sequence- and Function-Based Metagenomics

Extracting the Microbial Metagenome
• Tris boric acid EDTA (TBE)
• Low electroendosmosm (LE) agarose
• InstaGene matrix
• Polymerase chain reaction (PCR) tubes
• Triton X-100
• Phusion flash PCR master mix
• Nuclease-free water
• Loading dye
• Primers
• 100-bp low-scale DNA ladder
• Ethidium bromide
• Heat blocks (one at 100°C and one at 56°C)
• Thermal cycler
• Gel electrophoresis box and power source
• Imaging system with an ultraviolet (UV) transilluminator

Sequence- and Function-Based Metagenomics
• Luria-Bertani (LB)-agar with 7 μg/mL chloramphenicol
TABLE 1.
Alignment of lab modules, learning objectives, methods, and assessment questions.

<table>
<thead>
<tr>
<th>Lab Module</th>
<th>Module-Specific Learning Objectives/Outcomes</th>
<th>Skills/Methods</th>
<th>Assessment Questions (# and Topic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The Microbial World</td>
<td>1. Explain the risks and benefits of living with microbes</td>
<td>1. Sample collection</td>
<td>1. Order the key steps in a metagenomics experiment</td>
</tr>
<tr>
<td></td>
<td>2. Define metagenomics and explain its applications</td>
<td>2. Light microscopy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Postulate how the surrounding landscape influences the microbial populations in the Mississippi river</td>
<td>3. Microbe sizing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Outline the importance of microbes to river ecosystems</td>
<td>4. Microbe filtering</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. List three potential benefits of using metagenomics to study river microbes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Extracting the Microbial</td>
<td>1. Explain how and why the 16S rDNA differs among organisms</td>
<td>1. Extract genomic DNA</td>
<td>3. PCR</td>
</tr>
<tr>
<td>Metagenome</td>
<td>2. Analyze and interpret a rarefaction curve</td>
<td>2. Learn to use micropipettes and centrifuges</td>
<td>4. OTU</td>
</tr>
<tr>
<td></td>
<td>3. Explain how DNA storage and handling techniques minimize damage caused by metal ions and enzymes</td>
<td>3. Set up a PCR reaction</td>
<td>5. 16S rDNA</td>
</tr>
<tr>
<td></td>
<td>4. Illustrate the concept of an Operational Taxonomic Unit (OTU)</td>
<td>4. Formulate a hypothesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Predict antibiotic resistance profiles of Mississippi microbes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6. Diagram how PCR works at the molecular level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Sequence- and Function-Based</td>
<td>1. Compare and contrast sequence-based and function-based metagenomics and list the strengths and weaknesses of each</td>
<td>1. Perform gel electrophoresis</td>
<td>2. Sequence-based metagenomics</td>
</tr>
<tr>
<td>Analysis</td>
<td>2. Plan a functional metagenomics screen and explain how this technique can be used to uncover novel proteins</td>
<td>2. Analyze/interpret gel data</td>
<td>6. Gel electrophoresis</td>
</tr>
<tr>
<td></td>
<td>3. Explain how 16S rDNA analysis can classify microbe populations</td>
<td>3. Use sterile technique to plate E. coli clones</td>
<td>8. Table hypothesis</td>
</tr>
<tr>
<td></td>
<td>4. Define an E. coli fosmid library and explain how it can be used to identify proteins of interest</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Describe the principle behind gel electrophoresis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metagenomes</td>
<td>2. Explain how organisms are classified</td>
<td>2. Explore and use IMG</td>
<td>8. Table hypothesis</td>
</tr>
<tr>
<td></td>
<td>3. Interpret a phylogenetic tree</td>
<td>3. Perform comparative genomics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Identify homologous genes</td>
<td>4. Interpret a phylogenetic tree</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Identify homologs to commercially useful genes</td>
<td></td>
</tr>
</tbody>
</table>

PCR = polymerase chain reaction; IMG = Integrated Microbial Genomes website, https://img.jgi.doe.gov/cgi-bin/w/main.cgi.

- Escherichia coli fosmid libraries (control and test samples)
- Antibiotic discs
- Lazy-L-Spreaders (individually wrapped)
- Incubator (37°C)

Bioinformatic Analysis of Metagenomes
- Computer
- Internet access

Student instructions

Each of the four lab modules in the metagenomics series has a specific focus, yet they are also highly connected by a common metagenomics theme. Prior to the Microbial World lab module, each student collected an environmental water sample of his or her choice. During the Microbial World lab module, students 1) examined microbial diversity by observing and characterizing prokaryotic and eukaryotic microbes found in the commercially prepared pond water and recorded their findings in their lab book, and 2) collected the bacteria from their environmental water samples by filtration. The bacteria from those water samples were then frozen for use in the Extracting the Microbial Metagenome lab module, where students extracted the bacterial DNA. The extracted bacterial DNA served as a template in a modified polymerase chain reaction (PCR). Subsequently,
in the Sequence- and Function-Based Metagenomics lab module, the students (working in groups of four) synthesized much of what they learned in prior labs to design their own antibiotic-resistance experiment. These experiments made use of *E. coli* libraries from the M3P Project, which were derived from different sampling locations along the Mississippi river. Each library consists of individual *E. coli* clones that harbor a piece of environmental DNA contained within a fosmid (i.e., a large circular DNA vector). Finally, in the Bioinformatic Analysis of Metagenomes lab module, the students analyzed phylogenetic distance trees and compared genomes to metagenomes. Handouts containing detailed instructions and resources can be found in Appendix 2.

**Faculty instructions**

Our experience with these lab modules has given us several general insights into how to implement these labs successfully. First, as with any new lab, it is recommended that instructors and lab support personnel perform these labs themselves before testing them with students. Second, students will likely perform better if these modules are coupled with instructional videos (see Appendix 2) and/or lectures about metagenomics before entering the lab. Third, we have found that creating small aliquots of the reagents for the students helps to minimize contamination issues and results in a more uniform experience for the students. Fourth, students should be explicitly encouraged to think critically when they are designing their experiments and reminded that the best experiments usually involve additional research and group meetings outside of class. Finally, it is very important to keep the students on task; the time students have enough time to finish. More information about each lab module is available in Appendix 2. Additional detailed faculty instructions can be found in Appendix 3.

**Suggestions for determining student learning**

Assessments for monitoring student learning and evaluating student performance (grades) used formative and summative methods, respectively. Formative assessment was achieved through questions asked in the handouts. In the Microbial World lab module, students recorded the characteristics of the microbes they observed. In the Extracting the Microbial Metagenome and Sequence- and Function-Based Metagenomics lab modules, students analyzed and drew conclusions from experimental data (i.e., sample diversity and antibiotic resistance respectively). Also, in the Sequence- and Function-Based Metagenomics lab module, an informal presentation was given. In the Bioinformatic Analysis of Metagenomes lab module, the students performed specific bioinformatic analysis activities as directed in the handout and recorded their results. For each of the lab modules, short-answer questions were provided in the handout, which the students used to check their understanding of the material. Summative assessment for the purpose of assigning a grade was accomplished through prelab and major quizzes. Detailed descriptions of the assessment tools can be found in Appendix 4.

To determine student learning of metagenomics concepts and the effectiveness of the metagenomics lab modules, a pre-/posttest assessment was designed. This assessment consisted of eight multiple-choice content questions, four student information questions, and one control question (Appendix 5). We encouraged students to take the assessment by offering them bonus points for completion of the pre-/posttests. Students who scored less than 50% on these tests earned one bonus point, and students who scored greater than 50% earned two bonus points.

**Sample data**

During the Microbial World lab module, each student completed a table of the characteristics (i.e., a drawing, an estimate of the size, a description of the mode of motility, etc.) of the microbial organisms they found in commercially prepared pond water samples. In the Extracting the Microbial Metagenome lab module, students ran and analyzed their PCR product on an agarose gel. Up to 11 student samples could be observed on each gel image. During the Sequence- and Function-Based Metagenomics lab module, each group of four students examined three LB-agar plates with 7 μg/mL chloramphenicol and selected antibiotic discs to determine antibiotic resistance profiles. Each group recorded data on the size of the clearing zone around the antibiotic discs on the plate, the number and size of resistant colonies, and the distance of resistant colonies from the disc. Finally, in the Bioinformatic Analysis of Metagenomes lab module, the students used the Integrated Microbial Genome (IMG) website to analyze phylogenetic distance trees and compare genomes to metagenomes. Results of these analyses and comparisons were recorded in the handout. Examples of expected outcomes and student sample data are presented in Appendix 6.

**Safety issues**

Students used standard safety practices and wore standard laboratory protection. Other specific safety instructions are listed below.

The Microbial World
1. Watch your footing when collecting environmental water samples.
2. Collect and dispose of methyl blue and methyl green stains in accordance with the guidelines specified by the chemical hazardous waste program in your state and institution.
3. Thoroughly wash your hands before leaving the lab after examining live microbes.
4. Use a broom and dustpan when cleaning broken glass, not your hands (even if gloved).
Extraction the Microbial Metagenome

1. Do not use the power supply if the electrical outlets are wet.
2. Do not operate the electrophoresis gel box without a safety cover; it may result in a shock.
3. Do not touch the heat blocks as this could result in a burn.
4. Use a face-shield (or equivalent) and protective clothing when examining the gel. Ultraviolet light can damage the retinas of your eyes and cause skin burns.
5. Use gloves when handling the gels as they contain ethidium bromide, a mutagen and carcinogen.

Sequence- and Function-Based Metagenomics

1. All plates and equipment (e.g., spreaders and pipette tips) used with *E. coli* should be discarded in a biohazard container that has a lid and contains a clear biohazard bag. The biohazard bag is autoclaved when the lab is complete.
2. Closely monitor the forceps when they are being sterilized with the burning alcohol lamp to prevent fires.
3. Ask students with known allergies to specific antibiotics to not touch anything that has come in contact with those antibiotics (e.g., antibiotic discs, forceps, or plates).

Bioinformatic Analysis of Metagenomes

1. There are no safety issues associated with this activity.

Discussion

Field testing

These lab modules have been tested and revised for the past seven semesters and they have been used in their current form for the past two semesters. A total of 635 students completed these modules in the fall of 2013. These students were divided into 32 lab sections, which each had 20 ± 4 students. Student demographic data for those included in this study are shown in Table 2. The majority of these students were sophomores with undeclared majors. Ethnicity, gender, class standing, and international status data indicate a moderate level of student diversity in our study population.

These labs will likely scale well to smaller or larger groups. Set-up time is similar to other labs in this course. These labs are not cookie cutter labs; they provided students with an engaging research experience. However, they are more costly than the other labs in this course (Appendix I).

Feedback on these lab modules has been overwhelmingly positive. The teaching assistants indicated that the hands-on interactive nature of these modules keeps students engaged and invested in learning the final outcome. The non-biology majors in the University Honors Program who piloted these modules were also very impressed. We do not have space to include all of their laudatory comments but here is a representative sampling:

- *It gave me a really concrete example of some work that I could do, a career path.*
- *It made me interested in possibly pursuing a career in microbiology.*
- *The text and lab got me more interested in the area of functional metagenomics.*
- *Learning about how medicine can be derived from microbes seems like it would be interesting to research.*
- *I just really feel that the hands-on experience we get to do really helps me to understand our material more.*
- *The best part was examining the resistances of the microbes at certain locations and hypothesizing the reasons why.*

Evidence of student learning

Data from the pre-/posttests were processed prior to analysis. Data were eliminated from students who 1) did not elect to participate in the study, 2) did not take both the pretest and the posttest, or 3) did not answer our control question correctly. The control question, “The University of Minnesota is in which state?” was intended to identify students who were answering the questions randomly. A total of 504 out of 635 students met the inclusion criteria. The average percent correct was calculated for each question and for the test as a whole. A single-tailed paired t-test was performed to compare the pretest and posttest data for each question. *P* values less than 0.05 were regarded as significant.

The results of this pre-/posttest analysis showed that students performed significantly better on the posttest than they did on the pretest (Fig. 1). This is true for each individual question and for the test as a whole.
analyzing the pre-/posttests in their entirety, the results showed that student performance increased from an average of 49% correct on the pretest to an average of 71% correct on the posttest. The most significant learning gains were seen for questions 5 (47% increase) and 6 (48% increase), which related to 16S rRNA and gel electrophoresis, respectively (Fig. 1). Poor pre-/posttest performance on question 2 indicates that students struggled the most with sequence-based metagenomics (Fig. 1 and Appendix 5), an intrinsically complex topic that perhaps needed additional explanation by the teaching assistants.

Student performance on the pretest exceeded our expectations. More than half of the students were able to answer questions 1, 3, 4, 7, and 8 correctly (Fig. 1 and Appendix 5). Student scores were substantially higher than the 25% correct that would be expected due to random guessing alone. This result was very surprising given that 89% of respondents had no prior research lab experience and 97% indicated that they had not been exposed to metagenomics in any of their prior classes (data not shown). Despite high pretest scores on these questions, students increased their posttest performance by working through these metagenomics modules. For example, the percentage of students that answered question 1 correctly went from 78% on the pretest to 96% on the posttest ($p < 0.0001$).

The multidisciplinary nature of metagenomics can be used to introduce a wide range of different biology topics. Evolution, microbial ecology, phylogeny, and genetics are just some of the topics that are addressed with these modules. Learning gains on questions 3, 6, 7, and 8, which address PCR, gel electrophoresis, distance trees, and hypothesis formation, respectively, show that these modules can be used to teach core biology topics in addition to metagenomics (Fig. 1 and Appendix 5).

### Possible modifications

There are many ways in which these modules could be adapted to different course conditions. While these modules are designed to be taught together, they can also stand alone to address individual concepts. In addition, the timing of the modules could be modified. In order to align these modules with the lecture material, we elected to have several weeks between the Microbial World module and the subsequent three lab modules. An alternative approach would be to teach these lab modules consecutively to minimize the chance that students will forget material covered in the first lab before doing the subsequent metagenomics modules.

Instructors with more time flexibility and resources could have students perform DNA sequencing on their own metagenomics samples. The sequencing can be done in-house in institutions with high-throughput sequencers, or it can be outsourced. Care should be taken before adopting a sequencing approach; while it can make the metagenomics labs much more authentic, it also makes them more costly and more time-consuming as it usually takes up three to four weeks for a sample to be sequenced.

The functional metagenomics module involves working with a mixed *E. coli* fosmid library that was created by independently culturing isolated metagenomics clones and then subsequently mixing them together. One of the problems with this approach is that mixed libraries can be challenging
to amplify. If any individual E. coli clone has a growth advantage, it could quickly dominate the library. An alternative approach that allows for easier propagation of the library is to grow and test each clone individually. While this is a closer approximation to how metagenomics screens really work, students will be able to screen far fewer clones, and some students may not find resistance to their chosen antibiotics. Regardless of how the screen is performed, another simple modification would be to change which antibiotics are tested.

The bioinformatics module is the most flexible as it allows instructors to explore optional activities without incurring extra costs. One option for more advanced courses is to use the MOTHUR program (www.mothur.org/) to analyze the metagenomics sequence reads. This program is often used by researchers in the field, it is free, and it has a wealth of online documentation. Students doing a MOTHUR module can learn about how sequences are processed before they can be effectively analyzed. Alternatively, GENI-ACT (www.geni-act.com/) allows students to annotate genes in metagenomics samples, and it features online lab notebooks for easy grading.

SUPPLEMENTAL MATERIALS

Appendix 1: Materials and equipment
Appendix 2: Handouts for laboratory modules
Appendix 3: Additional faculty instructions
Appendix 4: Additional suggestions for determining student learning
Appendix 5: Metagenomics module pre-/posttest
Appendix 6: Student sample data

ACKNOWLEDGMENTS

We would like to thank all of the Biol 1009 students, instructors, TAs, and lab staff who participated in or contributed to these lab modules. In particular we would like to thank Cody Smith, our senior lab coordinator, for all his help with testing and running these modules. We also would like to thank Dr. Michael Sadowsky for allowing us to use the functional metagenomics library from the Minnesota Mississippi Metagenomics Project. This program was supported in part by a grant to the University of Minnesota-Twin Cities from the Howard Hughes Medical Institute through the Science Education Program. The authors declare that there are no conflicts of interest.

REFERENCES